Phonological neighbors in a small world:

What can graph theory tell us about word learning?

Michael S. Vitevitch

Department of Psychology

University of Kansas

Running Head: Phonological neighbors in a small world

Correspondence should be addressed to:
Michael S. Vitevitch, Ph.D.
Spoken Language Laboratory
Department of Psychology
1415 Jayhawk Blvd.
University of Kansas
Lawrence, KS 66045
e-mail: mvitevit@ku.edu
ph: 785-864-9312

ABSTRACT

Graph theory mathematically describes the general principles that govern the development and organization of complex systems. Several analyses examined phonological word-forms in the mental lexicons of adults and of children in the context of graph theory. The results show that the adult lexicon has a short path length, a high clustering coefficient, and a power-law degree distribution. These characteristics were not present at earlier points in time (16- and 18-months of age), suggesting that the lexicon undergoes significant restructuring and self-organization over time. The implications of viewing the lexicon from a graph theoretic perspective for word learning and lexical access during perception and production are discussed. This perspective may also unify accounts of the evolution, development, and processing of language, as well as connect Psychology and Cognitive Science to a more universal theory that underlies many complex systems found in the real world.

Many objects in the world and the relationships that exist between them can be represented graphically as a network with objects depicted as *nodes* (sometimes called vertices) and relationships between nodes depicted as *links* (sometimes called edges). For example, people in a city can be represented as nodes, with links connecting them if the people are acquainted with each other. Similarly, web pages on the World Wide Web may be represented as nodes, and hyperlinks connecting one page to another would constitute a link between nodes. The depiction of diverse problems from several different areas of research as abstract graphs (or networks) has enabled mathematicians in the sub-field known as *graph theory* (as well as physicists and computer scientists) to discover some of the general principles that govern complex systems in the real world.

Some systems are highly regular and can be depicted as a lattice (see the top panel in Figure 1). Notice that each node is linked to each adjoining node in the system, like a ring of people holding hands. In order to communicate a message from the node (i.e., person) labeled 1 to the node labeled 5, the message must go through nodes 2, 3, and 4. Similarly, to communicate a message from node 1 to node 6, the message must go through nodes 2, 3, 4 and 5. More links must be traversed if the message must go to a more distant node.

Insert Figure 1 about here

In contrast, some systems, as in the bottom panel of Figure 1, are connected at random (see Erdos & Rényi, 1960, for pioneering work on random graphs). Continuing with the example of trying to send a message between nodes, consider again sending a message from node 1 to node 6. In this case, a direct link exists, enabling rapid transmission of the message. Although fewer nodes were traversed to communicate the message from node 1 to node 6 in the random graph than in the lattice, this may not always be the case. If the links were distributed in a different random configuration, it may be impossible to ever communicate a message from node 1 to 6 (consider sending a message from node 1 to 4 in the present configuration). Now consider sending a message from node 1 to node 5. In this case, there is no direct link between the two nodes. In order to communicate the message it must be sent indirectly through nodes 9, 7, and 10. An example of a relatively random network in the real world is the national highway system: a highway directly links some, sometimes distant, cities. In the case of other cities, however, you can't get there (directly) from here.

Between the extremes of ordered and random graphs lies another class of graphs—collectively called *small world* networks—that has recently received a great deal of attention from graph theory researchers (e.g., Albert & Barabási, 2002; Watts & Strogatz, 1998). These graphs, which are representative of many complex systems in the real world, have some characteristics of regular graphs, and some characteristics of random graphs. The right combination of these characteristics enables these systems to be dynamic and to convey information very efficiently. Despite fragile appearances, these systems are robust and highly resistant to damage. The present analyses will examine the organization of phonological word-forms in the mental lexicon of adults to determine if these representations are structured in a regular fashion like a conventional dictionary, randomly listed in lexical memory, or organized in a way that lies somewhere in between these two extremes, making the lexicon an efficient, dynamic system that is resistant to error or damage.

What makes a small-world?

Small-world networks are identified by their characteristic path lengths and clustering coefficients (e.g., Barabási & Albert, 1999; Watts, 1999; Watts & Strogatz, 1998). *Path length* refers to the number of links that must be traversed to connect any two nodes in the network. The maximum path length, or longest path between two nodes in the network, is referred to as the *diameter* of the network. Despite being rather large systems (i.e., they have many nodes), a relatively small number of links must be traversed to reach any two nodes in the network as evidenced by a small average path length (and diameter). The relatively small distance between any two nodes in the graph affords the network great processing efficiency.

Nodes in small-world networks also tend to form clusters. A cluster can be thought of in the following way: "…your friend's friend is also your friend" (Newman, 2003; pg. 56). That is, several nodes connect to a common node, and to each other as well. The probability that two neighbors of a given node are connected to each other is referred to as the *clustering coefficient* (Watts & Strogatz, 1998). Small-world networks have a high clustering coefficient compared to the clustering coefficient of a randomly structured network of comparable size, indicating that the neighbors of a given node are highly interconnected. This characteristic adds to the efficiency of the network and contributes to the robustness and resilience of the network in the face of damage.

One type of small world graph is a *scale-free network* (Albert & Barabási, 2002; Barabási & Albert, 1999). Like other small-world networks, a scale-free network has a relatively small path length and a high clustering coefficient. A scale-free network is distinguished from other types of small-world graphs (i.e., networks that lie between ordered and random graphs) by its unique degree distribution. The number of links that a node has is referred to as *degree* (k). One can plot the number of nodes that have one link, the number of nodes that have two links, etc. on a frequency distribution. When plotted on a log-log scale, this frequency distribution is referred to as the *degree distribution*. Figure 2 displays two common types of degree distributions.

Insert Figure 2 about here

In random graphs, the links between nodes are randomly placed with equal probability producing a degree distribution that approximates a Poisson distribution (the top panel in Figure 2). That is, in a randomly connected network almost every node will have the same number of links, with only a small number of nodes having greater or fewer links. The peak in the Poisson distribution corresponds to the "stereotypical node" in the network, or the *scale* of the network. In contrast, scale-free networks do not have a "stereotypical node" in the degree distribution; therefore, they are free of a scale, or "scale-free." The degree distribution of scale-free networks does not resemble a Poisson distribution, but rather typically follows a power-law distribution as illustrated in the bottom panel of Figure 2 (see Dorogovtsev & Mendes, 2003 for other types of small-world graphs and the degree distribution that those graphs produce). Typically, the slope of the best fitting line of the degree distribution—the *degree exponent*, γ—in a scale-free network is $2 < \gamma < 3$ (cf., Montoya & Solé, 2002).

Barabási and Albert (1999; Barabási, Albert & Jeong, 1999) suggest that scale-free graphs with power-law degree distributions arise as the result of two mechanisms: *growth* and *preferential attachment.* Growth refers to the addition of new nodes to the network over time. Preferential attachment is the tendency for new nodes to connect to nodes that are already highly connected. Barabási and Albert (1999; Barabási et al., 1999) found that *both* growth and preferential attachment are necessary to create the power law degree distribution in scale-free networks. They suggested that a complex system that does not grow or that grows without preferential attachment does not exhibit a power-law degree distribution (cf., Ferrer i Cancho & Solé, 2001a, for other mechanisms that may account for a power-law degree distribution).

Albert and Barabási (2002) discussed many complex systems in the real world that exhibit the characteristics of scale-free networks, including the World Wide Web, the Internet, collaborations among movie actors and among scientists, human sexual contacts, cellular networks, ecological networks, and phone call networks. Given the ubiquity of scale-free networks in natural and artificial systems found in the real world, it is perhaps not surprising that the same principles might organize complex *cognitive* systems. Indeed, Yook, Jeong, and Barabási (2001 as cited in Albert & Barabási, 2002; see also Batagelj, Mrvar & Zaversˇnik, 2002; Ferrer i Cancho & Solé, 2001b, Motter et al., 2002, and Steyvers & Tenenbaum, submitted) demonstrated that semantic concepts—defined variously as synonym-pairs, core words from dictionary definitions, co-occurring words in text, or free-associates obtained from Nelson, McEvoy, and Schreiber (1999)—exhibit the characteristics of scale-free networks.

In the present analysis phonological representations of English words were examined to determine if word-forms in the mental lexicon of adults were organized with the same principles that govern the development of scale-free networks. By viewing phonological word-forms from a graph theoretic perspective, unique insight might be gained with regards to the development, learning, and processing of spoken language.

Specifically, how might the organization of word-forms in the lexicon influence the learning of novel phonological forms of new words (e.g., Beckman & Edwards, 2000; Gathercole, Hitch, Service, & Martin, 1997; Storkel, 2001, 2003; Storkel & Morrisette, 2002), and how does the organization of phonological word-forms in the lexicon influence lexical access during the perception and production of spoken language?

Graph theoretic analysis of phonological word-forms in the lexicon

Approximately 20,000 words (*n* = 19,340) from a database of computer-readable phonological transcriptions (Nusbaum, Pisoni, & Davis, 1984) were examined with Pajek, a program for large network analysis and visualization (Batagelj & Mvrar, 1998). In the graph of the adult lexicon, each node represented a word in the database. A link connected two nodes if the two words were phonologically similar. Phonological similarity was operationally defined with a single phoneme substitution, addition or deletion metric (Landauer & Streeter, 1973). For example, a link was placed between the nodes for the word *cat* and the words *cap, hat, cut, at* and *scat* because a single phoneme can be substituted, added or deleted from *cat* to form those words. This metric was selected because of its use in previous studies that examined the influence of phonological similarity on word learning in children (e.g., Charles-Luce & Luce, 1990, 1995; Storkel, 2002) and the influence of phonological similarity on language processing in adults (e.g., Luce & Pisoni, 1998; Vitevitch & Luce, 1999). Other methods to assess similarity among word-forms will produce comparable results (e.g., Batagelj et al., 2002; also note that comparable results were found among Ferrer i Cancho & Solé, 2001b;

Steyvers & Tenenbaum, submitted, etc. using several different metrics of semantic similarity).

Several characteristics of the lexical network must be assessed and compared to a comparably sized random network (i.e., Erdos & Rényi, 1960) to determine if the phonological word-forms in the adult lexicon are organized like a small-world network, and more specifically like a scale-free network. The following characteristics must be assessed: average path length, diameter, clustering coefficient, and degree exponent. These characteristics (except for the degree exponent) were assessed in the lexical network with the Pajek network program. The same program was used to generate ten randomly linked networks of comparable size for comparison.

The values for the average path length and diameter of the lexical network should be comparable to those values obtained from a randomly connected network. If the lexical network has a small-world topology, then the clustering coefficient for the lexical network should be much greater than the clustering coefficient for a comparably sized random network. The values of these characteristics for the phonological word-forms in the lexical network and the mean values of these characteristics for 10 comparably sized random networks are displayed in Table 1.

Insert Table 1 about here

The average path length for the lexical network was 6.05. That is, on average approximately 6 links had to be traversed to connect any two nodes in the network. For example, to get from the word *cat* to the word *dog*, one must traverse the links between the nodes corresponding to the words *bat*, *bag*, and *bog*. This value accords well with the mean value obtained from the random networks (8.44).

The maximum path length, or diameter, of the lexical graph was 29 links. That is, the longest path between two nodes in the graph was the 29 links between the word *connect* and the word *rehearsal* (via *collect, elect, affect, effect, infect, insect, inset, insert, inert, inurn, epergne, spurn, spin, sin, sieve, live, liver, lever, leva, leaven, heaven, haven, raven, riven, rivet, revert, reverse, rehearse*). Although the obtained value is somewhat larger that the mean diameter value obtained from the random networks (19.0), it is not greatly divergent (i.e., different by an order of magnitude).

Recall that the clustering coefficient measures the probability that two neighbors of a given node are connected to each other (Watts & Strogatz, 1998). Small-world networks have a larger clustering coefficient compared to the clustering coefficient of a randomly structured network of comparable size, indicating that the neighbors of a given node are highly interconnected. The phonological word-forms in the lexical network had a clustering coefficient of .045, which is over 250 times larger than the clustering coefficient for a comparably sized random network (.000162). The values for average path length, diameter, and the clustering coefficient obtained for the lexical network suggest that phonological word-forms in the adult lexicon are not stored as a random list of words, nor are they rigidly organized like a conventional dictionary. Rather, the organization of phonological word-forms in the mental lexicon of adults lies between randomness and complete order. That is, phonological word-forms in the mental lexicon exhibit a small-world topology.

To determine if the mental lexicon of adults is a scale-free network the degree exponent must be assessed. Recall that scale-free networks exhibit a degree distribution that approximates a power-law with a degree exponent, γ, of $2 < \gamma < 3$. Figure 3 displays a frequency distribution of the lexical network with the number of links per node on the x-axis, and the number of nodes that have a given number of links on the y-axis. The inset in Figure 3 displays this information on a log-log scale (i.e., the degree distribution).

Insert Figure 3 about here

In a randomly connected network the degree distribution follows a Poisson distribution, however, in a scale-free network the degree distribution follows a power-law (see Figure 2). As shown in Figure 3, the phonological word-forms in the mental lexicon exhibit a degree distribution that approximates a power-law. The degree exponent, γ, was 1.96, which closely approximates the values of $2 < \gamma < 3$ typically observed for scale-free networks. The results of the present analyses suggest that the phonological word-forms in the mental lexicon of adults are, indeed, organized in a scale-free topology.

The lexicon as a scale-free network

Recall that Barabási and Albert (1999; Barabási et al., 1999) suggested that the power-law degree distribution that is characteristic of scale-free topologies is the result of growth and preferential attachment in the system. Although learning new words is something that is typically associated with and primarily studied in children (e.g., Storkel 2001, 2003), adults can and do learn novel sound patterns/new words (e.g., Storkel, Armbrüster & Hogan, submitted), suggesting that even the adult lexicon undergoes growth.

The growth that occurs in scale-free networks does not proceed randomly, but instead occurs with preferential attachment. That is, a new node being added to the system will tend to connect to a node in the system that is already highly connected rather than connect to a sparsely connected node in the system. If preferential attachment played a role in organizing the phonological word-forms in the mental lexicon of adults, then a word that was learned early in life should be linked to many similar words, whereas a word that was learned later in life should be linked to few similar words. Indeed, Storkel (2004) found a positive correlation between age-of-acquisition and *neighborhood density* or, the number of words that are phonologically related to a target word. Words acquired early in life tended to have many phonological neighbors (i.e., dense phonological neighborhoods), whereas words acquired later in life had fewer phonologically neighbors (i.e., sparse phonological neighborhoods). If preferential attachment plays a role in organizing the mental lexicon, one would further predict that phonological neighborhoods become "denser" over time. Charles-Luce and Luce (1990, 1995) analyzed a set of words in the lexicons of adults and children 5- and 7-years old, and found that the neighborhood density for words in the adult lexicon were indeed denser than the neighborhood density for those same words in the 5- and 7-years old lexicons. These findings are consistent with the hypothesis that preferential attachment guides growth in the mental lexicon and other complex systems with scale-free topologies.

Another hypothesis that one might derive from the mechanism of preferential attachment is that words with dense phonological neighborhoods should be more easily acquired or learned than words with sparse phonological neighborhoods. Storkel (2001, 2003) found that pre-school age children learned novel words that had common sound sequences/dense neighborhoods more rapidly than novel words that had rare sound sequences/sparse neighborhoods. Similarly, Storkel et al. (submitted) found that college-age adults learned novel words with dense neighborhoods more rapidly than novel words that had sparse neighborhoods. These results suggest that the learning of novel words is highly dependent on the words that are already known. Specifically, a novel word that is similar to many known words may be learned more quickly than a novel word that is similar to few known words. Moreover, this set of seemingly unrelated results is consistent with the predictions derived from preferential attachment, a mechanism that plays a role in the growth and development of scale-free networks (e.g., Barabási & Albert, 1999).

Other characteristics of complex systems that exhibit scale-free topologies include efficient processing, and robustness to error or damage. For example, Albert et al., (2000) found in their analysis of the World Wide Web that damage to a scale-free network does not result in catastrophic failure in the system. Rather, damage in scale-free networks tends to affect the less connected nodes because (as shown in Figure 3) there is a greater prevalence of such nodes in the system. Although there are few nodes that are highly interconnected (highly interconnected nodes are sometimes referred to as *hubs*), the interconnectedness of these nodes maintains the integrity of the whole system. Even if a hub is damaged, the presence of other (but relatively less connected) hubs will absorb the extra load and enable processing to continue. Only if every node has been damaged or removed will a scale-free network catastrophically fail.

When viewing the mental lexicon as a complex system with a scale-free topology, we also see quite rapid and relatively error-free language processing in normal adults (e.g., Levelt, 1989). In addition, we observe patterns of topological robustness in the mental lexicon that resemble the pattern found by Albert et al., (2000) in their analysis of the World Wide Web: error/damage tends to affect the less connected nodes. For example, Vitevitch (1997; 2002a) found that normal adults made more speech production errors for words with few phonological neighbors than for words with many phonological neighbors (see also Harley & Bown, 1998; James & Burke, 2000; Vitevitch & Sommers, 2003). More errors for words with few phonological neighbors than for words with many phonological neighbors were also observed in experiments investigating short-term memory (e.g., Roodenrys et al., 2002), and in patients with aphasia (e.g., Gordon & Dell, 2001). Finally, words with few phonological neighbors were also produced more slowly than words with many phonological neighbors (Vitevitch, 2002a). These findings are consistent with the hypothesis that the mental lexicon is a (topologically robust) scale-free network, and that error or damage in such systems primarily affects the less connected nodes.

Considering the phonological word-forms in the adult mental lexicon as a complex system with a scale-free topology produced several predictions that were not only consistent with findings from studies of word learning in children and adults, but also produced predictions that were consistent with findings from studies of adult language processing. There are few (if any) models of adult language processing that also account for the growth and development of that system (e.g., McClelland & Elman, 1986). Similarly, theories of word learning account for how new word-forms are added to the lexicon, but say little (if anything) about the changes in processing that might result from the addition of that new word-form to the lexicon (e.g., Gupta & MacWhinney, 1997). By taking a graph theoretic perspective of the mental lexicon, it may be possible to describe the mechanisms that underlie the acquisition and learning of novel word-forms, as well as the processes involved in on-line language processing in one theoretical account.

When does a scale-free lexicon emerge?

The previous analysis demonstrated that the organization of phonological word-forms in the adult lexicon exhibited a scale-free topology. The previous analysis, however, did not provide any information about *when* this highly efficient and robust lexical structure emerged during language development. A great deal of word learning research suggests that the mental lexicon may undergo restructuring or reorganization with the acquisition of new words over time (e.g., Charles-Luce & Luce, 1990, 1995; Strokel, 2002). The traditional—though not uncontroversial—benchmark for the onset of some form of cognitive reorganization that results in a “vocabulary spurt” in children occurs when 50 words have been acquired (e.g., Goldfield & Reznick, 1996; Mervis & Bertrand, 1995). Various mechanisms have been proposed for the vocabulary spurt (e.g., Golinkoff et al. 2000; Nazzi & Bertoncini, 2003). The present analyses will determine if this “critical period” is also the point at which a scale-free topology emerges in the lexicon.

The same network statistics calculated for the adult lexicon were calculated on the lexicons of children of various ages to better determine when a scale-free topology emerged from the lexicon. The norms from the MacArthur Communicative Development Inventory (CDI) were used to (conservatively) estimate the lexicons of children at various ages. The earliest age at which 50% of the children knew a given word was used to determine which words children knew at a particular age. Network analyses were performed with Pajek for the lexicons of children at the end points of the CDI (16 months, n = 24 words, and 30 months, n = 490 words), as well as the two ages that straddled the 50-word point (18 months, n = 38 words, and 19 months, n = 78 words). The results of these analyses are shown in Table 2.

Insert Table 2 about here

Consider the network characteristics for the lexicon at 16- and 18-months of age. The average path length (ℓ), diameter (D), and clustering coefficient (C) for these graphs suggest that the lexical network is a small-world network. However, the degree exponent, γ, is considerably less than the values typically associated with scale-free networks ($2 < \gamma < 3$), which suggests that early in development, the mental lexicon may not have a scale-free topology (although it does lie somewhere between order and randomness). Not until 19-months of age does the degree exponent ($\gamma = 2.41$) reach a value typically associated with scale-free networks ($2 < \gamma < 3$); a similar value ($\gamma = 2.09$) was also observed at 30-months.

Recall that 18- to 19-months of age corresponds with the developmental milestone of "50-words" and the age at which the onset of the "vocabulary spurt" is often observed. It is striking that the present graph theoretic analyses demonstrated the emergence of a highly efficient and robust (i.e., scale-free) topology in the lexical network at the same age/developmental milestone. The results of this analysis support the idea that the phonological word-forms in the mental lexicon may indeed undergo some form of reorganization at this age/point in development. Once the phonological word-forms have self-organized into a scale-free network, the mechanisms of growth and preferential attachment may produce a complex cognitive system that is a highly efficient word learning and language processing machine. The variability in age/vocabulary size associated with reaching this developmental milestone may be due to different initial starting states of the system. That is, the first few sound patterns that are learned (e.g., Mandel, Jusczyk & Pisoni, 1995) may play a large role in determining how easily subsequent words are acquired.

Limitations of the present analyses

The present examination of the mental lexicon used a *minimal* scale-free model characterized by undirected and binary links between nodes, and a simple preferential attachment rule. Although viewing the mental lexicon within the context of this very simple model is consistent with several findings in diverse areas of language research, future examinations of the lexicon from the graph theoretic perspective may benefit from the use of different kinds of links and additional attachment-related parameters to better capture other findings in the literature.

Consider first the links used in the present analysis. If two words differed by a single phoneme, a link was assumed to exist between the word nodes. If two words differed by more than one phoneme, no link existed between the words. Although this simple metric accounts for much of the variability observed in spoken language experiments (e.g., Luce & Pisoni, 1998), it may not adequately capture the phonological similarity that might exist between words that differ by more than one phoneme, such as morphologically related words (e.g., *cover-uncover*), or smaller words that are embedded in longer words (e.g., *cat-catfish*). Instead of using binary links (absent or present) as in the present analysis, future graph theoretic accounts of the lexicon could weight the links with values that range from 0 to 1 to better capture similarity relationships among words with the same onset (Vitevitch, 2002b), among words that differ by more than one phoneme, or among phonological word-forms that are also orthographically similar (e.g., Ziegler, Muneaux & Grainger, 2003).

The links in the present analysis were also undirected. The implication of undirected links is that the similarity relationship between two words is symmetrical, which may not be a valid assumption (e.g., Tversky & Kahneman, 1981). Consider the case of small words embedded in longer words: *cat* may be judged to be more phonologically similar to *catfish* than *catfish* is to *cat*. The use of directed links in future graph theoretic accounts of the lexicon might better account for such asymmetric similarity relationships.

Consider now the preferential attachment mechanism described in the present analysis. A logical (though extreme) prediction would be that every word that an individual learned would resemble the first word that was acquired. Although the initial state of a system may influence development, there is no language in which *every* word sounds similar to *every other* word; comprehension under less than ideal conditions (e.g., noise) would prove most difficult in such a language. Albert and Barabási (2002) discussed other mechanisms that together might better account for the scale-free topologies observed in the real world. An example of one of these mechanisms is a *fitness* parameter, η (see for example Dorogovtsev & Mendes, 2000 for an age-related parameter). By varying the fitness of each node even a recently added (and therefore less connected) node might successfully compete with older more connected nodes to attract the links of newly added nodes.

In the case of the mental lexicon, the fitness parameter might be equivalent to occurrence-related characteristics of words, such as how often the word occurs in the language (word frequency), or how long ago the word last occurred (i.e., recency; MacKay & Burke, 1990); each of these factors influences how quickly and accurately a word is processed. Including additional parameters in future graph theoretic accounts of the lexicon might lead to better accounts of word learning and word retrieval processes, as well as better explanations of the changes that occur in those processes over time (i.e., as we age, or decrease the rate at which we learn new words once we have acquired a lexicon of sufficient size to function as an adult; see also Dorogovtsev & Mendes, 2003).

Advantages of the graph theoretic perspective

In viewing the mental lexicon as a complex system, we may no longer need to postulate separate accounts for acquisition/learning and adult processing. Previous accounts of word learning did not account for adult language processing (e.g., Gupta & MacWhinney, 1997), and previous accounts of adult language processing did not account for the learning of new words (e.g., McClelland & Elman). In contrast, the unique characteristics of scale-free networks do not exist without the complex interaction between the current topology of the system and the growth of the system via preferential attachment (Albert & Barabási, 2002). The interdependence of these factors in complex systems provides a unique opportunity to integrate previously disparate psychological theories describing word learning (e.g., Metsala & Walley, 1998; Storkel, 2002), adult language processing (e.g., Luce & Pisoni, 1998), and the effects of aging on language processing (e.g., Sommers, 1996).

Looking at the mental lexicon in graph theoretic terms places the lexicon, and other complex psychological systems, in a broader context, and allows us to see how cognitive systems—like other complex systems found in the world—are governed by the same underlying principles. Thus, accounts of language acquisition, word learning, and adult language processing may not need to be accounted for by mechanisms that are specific to language. Instead, the processes used to acquire and retrieve information from the lexicon can be considered special instances of universal processes found in all complex systems.

In addition to providing a way to integrate word learning in children and language processing in adults, a graph theoretic perspective might also provide some insight into a curious set of results from studies that examined language processing in adults. In English, Vitevitch (1997, 2002a) found that phonological neighbors facilitate the retrieval of word-forms during speech production. This result contrasts with what is typically found in studies of spoken word recognition, where phonological neighbors compete among each other during retrieval (e.g., Luce & Pisoni, 1998). The contradictory influence of phonological neighbors in speech production and word recognition in English is even more perplexing when one considers the findings of Vitevitch and Rodríguez (submitted): phonological neighbors facilitated word recognition in Spanish. What are the conditions in which phonological neighbors facilitate retrieval and what are the conditions in which phonological neighbors compete among each other during retrieval? Graph theoretic studies examining the situations that lead to cooperation or competition among agents in games such as the prisoner’s dilemma (e.g., Cohen, Riolo &

Axelrod, 1998) may provide some insight into the situations that lead to facilitation or competition among phonological word-forms.

The present graph theoretic analyses showed that phonological word-forms in the mental lexicon self-organize to form a scale-free network, typified by a small path length between nodes, a high clustering coefficient, and a power-law degree-distribution. The similarity of the mental lexicon to other complex systems in the real world suggests that universal principles may govern growth and development in all of these domains. Graph theory may represent a new paradigm for psychology that could not only unify and integrate previously disparate areas of research, but more importantly, may increase our understanding of complex cognitive systems by enabling us to examine the entire system rather than focus on one part of it.

Acknowledgements

This work was supported by Research Grant DC 04259 from the National Institute on Deafness and Other Communicative Disorders, National Institutes of Health. I would like to thank Steven B. Chin, John Colombo, Ed Auer, Mark Steyvers, and Holly Storkel for helpful comments, suggestions, and discussions.

References

Albert, R. & Barabási, A.L. (2002). Statistical mechanics of complex networks. *Review of Modern Physics, 74*, 47-97.

Barabási, A. L. & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*, 509-512.

Barabási, A.L., Albert, R. & Jeong, H. (1999). Mean-field theory for scale-free random networks. *Physica A, 272*, 173-187.

Batagelj, V. & Mrvar, A. (1998). Pajek-A program for large network analysis. *Connections, 21*, 47-57.

Batagelj, V., Mrvar, A. and Zaversˇnik, M. (2002): Network Analysis of Texts. Proceedings of the 5th Internatonal Multi-Conference Informatin Society - Language Technologies. Ljubljana, October 14 - 15, 2002. B, 143-148.

Beckman, M.E. & Edwards, J. (2000). The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Development*, *71*, 240-249.

Charles-Luce, J. & Luce, P.A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language, 17*, 205-215.

Charles-Luce, J. & Luce, P.A. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language, 22*, 727-735.

Cohen, M.D., Riolo, R.L., & Axelrod, R. (1998). *The emergence of social organization in the prisoner's dilemma: How context-preservation and other factors promote cooperation.* Santa Fe Institute Working Paper 99-01-002.

Dorogovtsev, S.N. & Mendes, J.F.F. (2000). Evolution of networks with aging of sites. *Physical Review E*, *62*, 1842-1845.

Dorogovtsev, S.N. & Mendes, J.F.F. (2003). Accelerated growth of networks. In S. Bornholdt & H.G. Schuster (eds.) *Handbook of Graphs and Networks: From the genome to the Internet*. Wiley-VCH. (pp. 318-341).

Erdos, P. and Rényi, A. (1960). On the Evolution of Random Graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, *5*, 17-61.

Ferrer i Cancho, R. & Solé, R.V. (2001a). *Optimization in complex netowrks*. Santa Fe Institute Working Paper 01-11-068.

Ferrer i Cancho, R. & Solé, R.V. (2001b). *The small world of human language*. Santa Fe Institute Working Paper 01-03-016.

Gathercole, S.E., Hitch, G.J., Service, E. & Martin, A.J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology, 33*, 966-979.

Goldfield, B.A. & Reznick, J.S. (1996). Measuring the vocabulary spurt: A reply to Mervis & Bertrand. *Journal of Child Language, 23*, 241-246.

Golinkoff, R., Hirsh-Pasek, K., Bloom, L., Smith, L.B., Woodward, A.L., Akhtar, N., Tomasello, M. & Hollich, G. (2000). *Becoming a word learner: A debate on lexical acquisition*. London, Oxford University Press.

Gordon, J.K. & Dell, G.S. (2001). Phonological neighbourhood effects: Evidence from aphasia and connectionist modeling. *Brain and Language*, *79*, 21-23.

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: computational and neural bases. *Brain and Language*, *59*, 267-333.

Harley, T.A. & Bown, H.E. (1998). What causes a tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. *British Journal of Psychology, 89*, 151-174.

James, L.E. & Burke, D.M. (2000). Phonological priming effects on word retrieval and tip-of-the-tongue experiences in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1378-1391.

Landauer, T.K. & Streeter, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior, 12*, 119-131.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, Mass: MIT Press.

Luce, P.A., & Pisoni, D.B. (1998) Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*, 1-36.

MacKay, D.G. & Burke, D.M. (1990). Cognition and aging: A theory of new learning and the use of old connections. In T.M. Hess (ed.) *Aging and Cognition: Knowledge Organization and Utilization*. (pp. 213-263) New York: Elsevier Science Publishers.

Mandel, D.R., Jusczyk, P.W. & Pisoni, D.B. (1995). Infants' recognition of the sound patters of their own names. *Psychological Science, 6*, 314-316.

McClelland, J.L. & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86.

Mervis, C.B. & Bertrand, J. (1995). Early lexical acquisition and the vocabulary spurt: A response to Goldfield & Reznick. *Journal of Child Language, 22*, 461-468.

Metsala, J.L. & Walley, A.C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. Metsala & L. Ehri (Eds.), *Word Recognition in Beginning Literacy* (Vol.4, pp. 89-120). Mahwah, NJ: Lawrence Erlbaum Associates.

Montoya, J.M. & Solé, R.V. (2002). Small world patterns in food webs. *Journal of Theoretical Biology, 214*, 405-412.

Motter, A.E., de Moura, A.P.S., Lai, Y.C., & Dasgupta, P. (2002). Topology of the conceptual network of language. *Physical Review E*, 65, xx.

Nazzi, T. & Bertoncini, J. (2003). Before and after the vocabulary spurt: Two modes of word acquisition? *Developmental Science, 6*, 136-142.

Nelson, D.L., McEvoy, C.L., & Schreiber, T.A. (1999). The University of South Florida Free Association Norms. http://www.usf.edu/FreeAssociation.

Newman, M.E.J. (2003). Random graphs as models of networks. In S. Bornholdt & H.G. Schuster (eds.) *Handbook of Graphs and Networks: From the genome to the Internet*. Wiley-VCH. (pp. 35-68).

Nusbaum, H.C., Pisoni, D.B. & Davis, C.K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. In *Research on Speech Perception Progress Report No. 10*. Speech Research Laboratory, Psychology Department, Indiana University, Bloomington, Indiana. (pp. 357-376).

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L.M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory & Cognition, 28*, 1019-1034.

Sommers, M.S. (1996). The structural organization of the mental lexicon and its contribution to age-related declines in spoken word recognition. *Psychology & Aging, 11*, 333-341.

Steyvers, M. & Tenenbaum, J.B. (2003). *The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth.* Manuscript submitted for publication.

Storkel, H.L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44,* 1321-1337.

Storkel, H.L. (2002). Restructuring of similarity neighborhoods in the developing mental lexicon. *Journal of Child Language, 29*, 251-274.

Storkel, H.L. (2003). Learning new words II: Phonotactic probability in verb learning. *Journal of Speech, Language, and Hearing Research, 46,* 1312-1323.

Storkel, H.L. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics, 25*, 201-221.

Storkel, H.L., Armbrüster, J. & Hogan, T.P. (submitted). *Forming new representations: The influence of known sounds and words on adult word learning*. Manuscript submitted for publication.

Storkel, H.L. & Morrisette, M.L. (2002). The lexicon and phonology: Interactions in language acquisition. *Language, Speech, and Hearing Services in Schools, 33*, 24-37.

Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453-458.

Vitevitch, M.S. (1997) The neighborhood characteristics of malapropisms. *Language and Speech, 40*, 211-228.

Vitevitch, M.S. (2002a). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory & Cognition, 28*, 735-747.

Vitevitch, M.S. (2002b). Influence of onset density on spoken-word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 270-278.

Vitevitch, M.S. & Luce, P.A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40,* 374-408.

Vitevitch, M.S. & Rodríguez, E. (submitted). *Neighborhood density effects in spoken word recognition in Spanish*. Manuscript submitted for publication.

Vitevitch, M.S. & Sommers, M.S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory and Cognition, 31,* 491-504.

Watts, D.J. (1999). *Small Worlds: The dynamics of networks between order and randomness. Princeton Studies in Complexity*. Princeton University Press, Princeton.

Watts, D.J. & Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393,* 440-442.

Ziegler, J.C., Muneaux, M. & Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory & Language, 48,* 779-793.

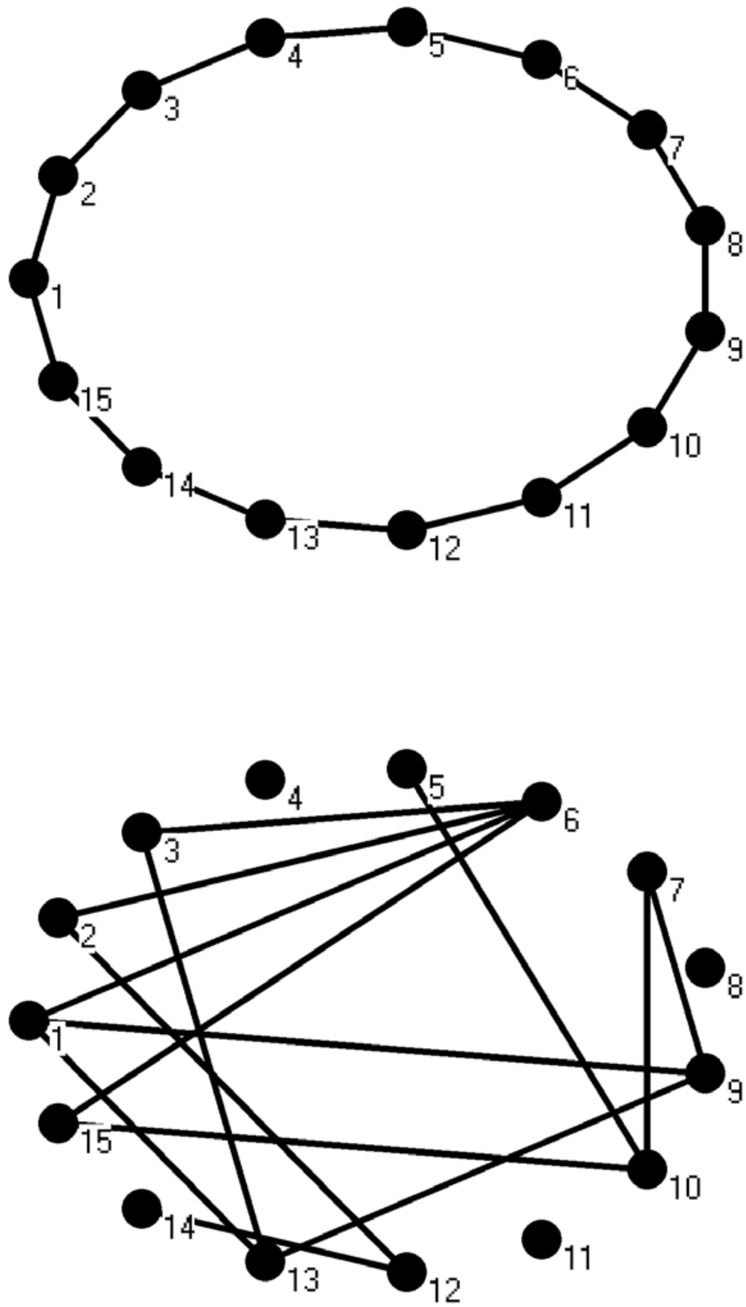


Figure 1. The top panel illustrates an ordered graph with 15 nodes and 2 links per node. The bottom panel illustrates a randomly connected graph with 15 nodes and (on average) 2 links per node. Graphs made with Pajek software (Batagelj & Mrvar, 1998).

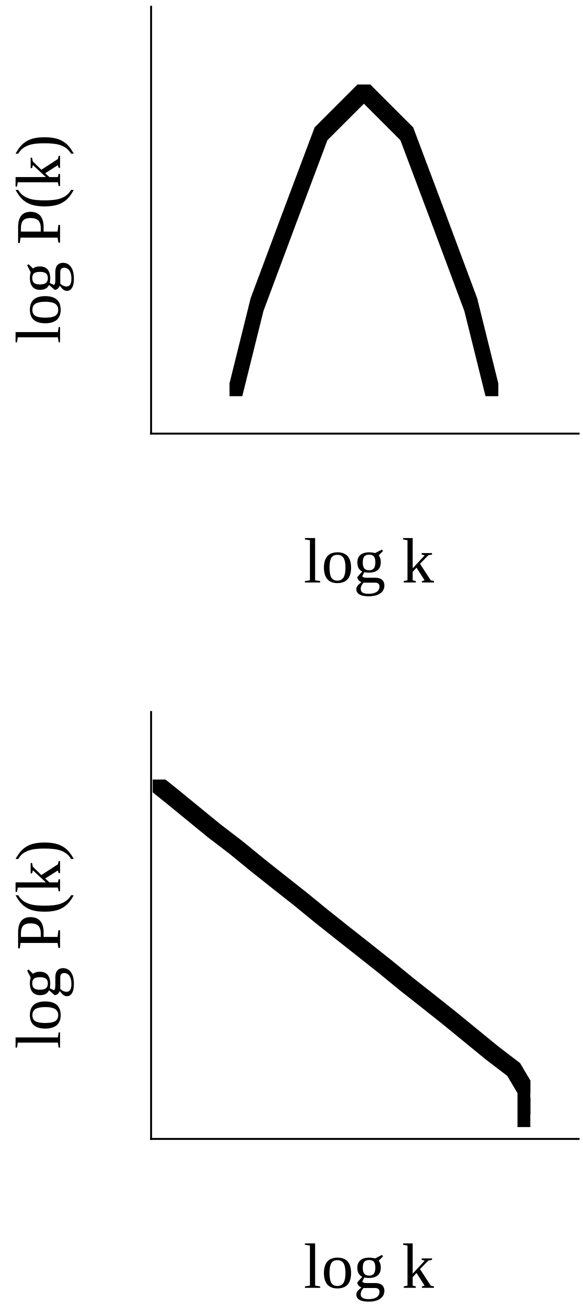
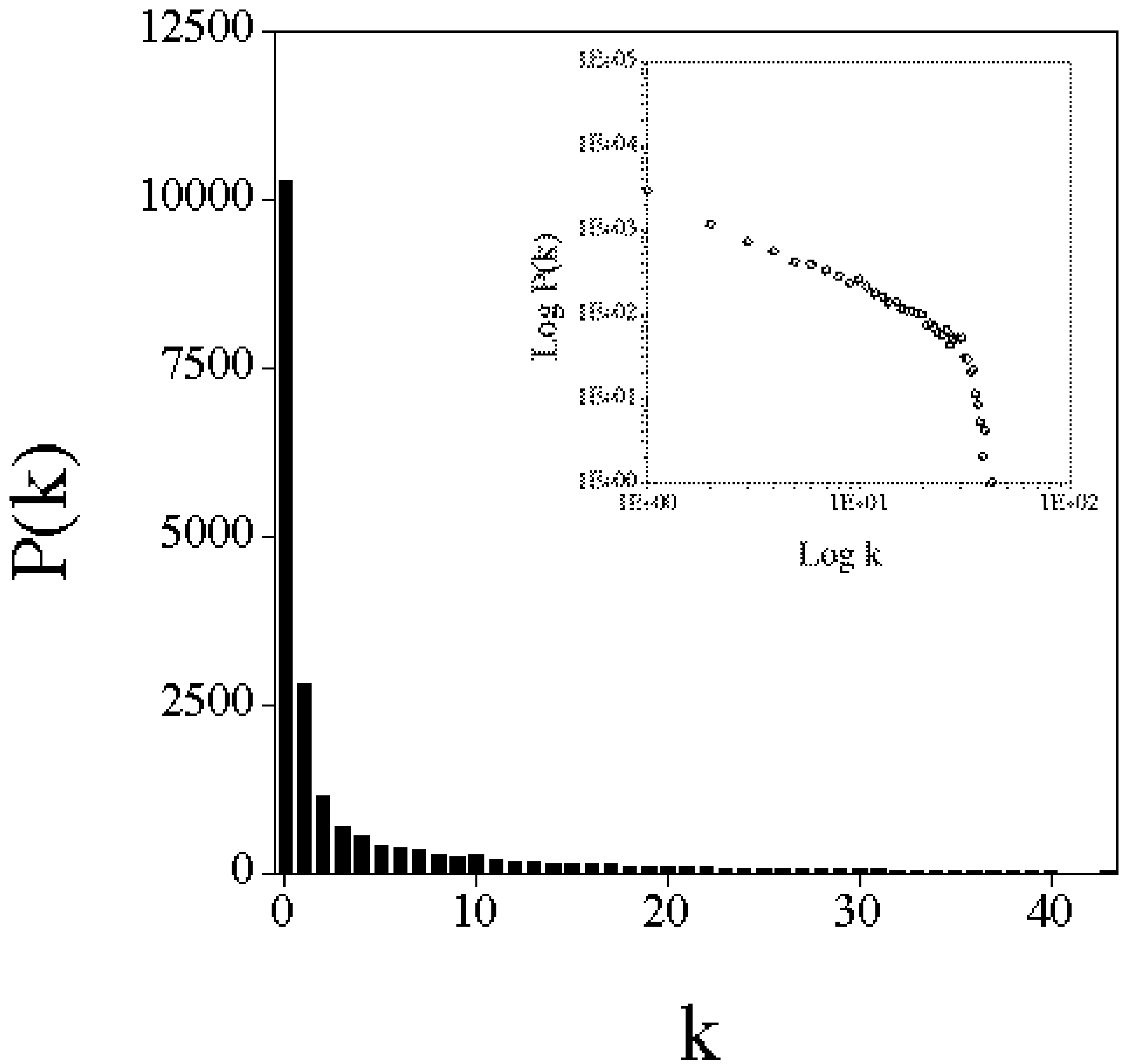


Figure 2. Examples of typical degree distributions in log-log plots. The top panel illustrates a Poisson distribution found in random graphs. The bottom panel illustrates a power-law degree distribution found in scale-free networks.



<u>Figure 3</u>. The number of nodes with *k* links (P*(k)*) as a function of number of links (*k*). The inset shows the same values plotted on a log-log scale (i.e., the degree distribution). Note the degree distribution follows a power-law.

Table 1. Network statistics for the adult lexicon and comparable random graphs

Network Characteristic	Lexical Network	Random Network
n	19,340	19,340
$\langle k \rangle$	3.23	3.23
ℓ	6.05	8.44 (.04)
D	29	19 (.816)
C	.045	.000162 (.000047)
γ	1.96	n/a

Note: For the random network the listed value is the mean (and standard deviation) from 10 simulations. n = number of nodes, $\langle k \rangle$ = average number of connections, ℓ = the average path length, D = the diameter of the network, C = the clustering coefficient, γ = degree exponent; typically $2 < \gamma < 3$ for scale-free networks.

Table 2. Network statistics for the lexicons of children at 16, 18, 19, & 30 months of age

Network Characteristic	16 months	18 months	19 months	30 months
n	24	38	78	490
$<k>$	.42	.39	.47	1.31
ℓ	1.00 [1.33]	1.11 [1.30]	1.65 [1.79]	6.28 [11.05]
D	1 [2]	2 [2]	3 [4]	17 [29]
C	.042 [0]	.026 [0]	.034 [0]	.089 [0]
γ	1.63	1.76	2.41	2.09

Note: n = number of nodes, $<k>$ = average number of connections, ℓ = the average path length, D = the diameter of the network, C = the clustering coefficient, γ = degree exponent; typically $2 < \gamma < 3$ for scale-free networks. For comparison, square brackets contain the values for ℓ, D and C for a comparably sized random network at each age. There were no clusters in any of the random graphs at any of the ages examined, therefore the values of C for the random graphs at all ages = 0.